https://doi.org/10.17221/119/2023-SWR

# Digital soil mapping using machine learning-based methods to predict soil organic carbon in two different districts in the Czech Republic

*Shahin Nozari[1]*, Mohammad Reza Pahlavan-Rad[2], Colby Brungard[3], Brandon Heung[4], Luboš Borůvka[1]*

[1]*Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic*
[2]*Soil and Water Research Department, Golestan Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Gorgan, Iran*
[3]*Department of Plant and Environmental Sciences, College of Agricultural, Consumer, and Environmental Sciences, New Mexico State University, Las Cruces, USA*
[4]*Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Truro, Canada*

**Corresponding author: nozari@af.czu.cz*

**Citation:** Nozari S., Pahlavan-Rad M.R., Brungard C., Heung B., Borůvka L. (2024): Digital soil mapping using machine learning-based methods to predict soil organic carbon in two different districts in the Czech Republic. Soil & Water Res., 19: 32–49.

**Abstract:** Soil organic carbon (SOC) is an important soil characteristic as well as a way how to mitigate climate change. Information on its content and spatial distribution is thus crucial. Digital soil mapping (DSM) is a suitable way to evaluate spatial distribution of soil properties thanks to its ability to obtain accurate information about soil. This research aims to apply machine learning algorithms using various environmental covariates to generate digital SOC maps for mineral topsoils in the Liberec and Domažlice districts, located in the Czech Republic. The soil class, land cover, and geology maps as well as terrain covariates extracted from the digital elevation model and remote sensing data were used as covariates in modelling. The spatial distribution of SOC was predicted based on its relationships with covariates using random forest (RF), cubist, and quantile random forest (QRF) models. Results of the RF model showed that land cover (vegetation) and elevation were the most important environmental variables in the SOC prediction in both districts. The RF had better efficiency and accuracy than the cubist and QRF to predict SOC in both districts. The greatest $R^2$ value (0.63) was observed in the Domažlice district using the RF model. However, cubist and QRF showed appropriate performance in both districts, too.

**Keywords:** cubist; DSM; quantile random forest; random forest; SOC

Although human population growth affects soil, soil quality must be maintained to ensure human survival (Pieri 1992; Brevik 2013). Soil organic carbon (SOC) is one of the most important indicators of soil quality and constitutes the largest terrestrial pool of bound carbon (Victoria et al. 2012; Lal et al.

Supported by the Czech University of Life Sciences Prague, Prague (internal Grant No. SV20-5-21130). The support from the Technology Agency of the Czech Republic, Project No. SS06010148, and the Ministry of Agriculture of the Czech Republic, Project No. QK22020217 is also acknowledged.

https://doi.org/10.17221/119/2023-SWR

2021). Many studies have been conducted to identify suitable methods to model and monitor SOC due to its substantial influence on atmospheric carbon dioxide ($CO_2$), which affects climate change (Selvaradjou et al. 2007). However, a high-resolution spatial prediction of SOC is needed to inform sustainable soil management practices and to assess the impacts of land-use.

Numerous studies have been conducted on the prediction of SOC distribution using digital soil mapping (DSM) (Nikou & Tziachris 2022); nevertheless, detailed aspects such as specific features, input data, and models used for spatial prediction in DSM have not been fully compiled for SOC in forest and agricultural soil (Minasny & McBratney 2016). Similarly in the Czech Republic, there are examples of producing high-resolution maps of SOC at the local, regional, or national scale, while there are no adequate studies considering the prediction of SOC in the Liberec and Domažlice districts. Additionally, there is no feasible study elucidating this approach, despite the region's active engagement in agriculture production. In this research, many different aspects of SOC are spatially evaluated concerning specific characteristics, input data, and models for SOC.

Therefore, this study aims to compare three models including random forest (RF), cubist, and quantile random forest (QRF) to assess their prediction accuracy, important variables, and spatial predictions of SOC as well as compare prediction uncertainty maps and suggest the best model that can be used to predict SOC in the Liberec and Domažlice districts in the Czech Republic.

## Background

This section reviews information about SOC distribution as well as applications of DSM and machine learning in the prediction of SOC distribution.

**SOC distribution.** One of the significant parameters that influence SOC distribution and explain the variation in SOC is topography as it is related to the extent of soil erosion, sediment yield, and the rate of incoming solar radiation. In addition, changes in other soil properties (such as changes caused by cultivation) influence the SOC content prediction by affecting aggregate stability, porosity, and bulk density. It has been also found that land-use, land management, vegetation, elevation, slope, rainfall, soil type, and wetness index are the most effective predictors of SOC (Badia et al. 2016; Mosleh et al. 2016; Wiesmeier et al. 2019; Borůvka et al. 2022). Additionally, Nozari & Borůvka (2020, 2023) showed that there is a clear relationship between SOC and environmental variables, particularly terrain parameters. Although changes in environmental variables can influence the SOC prediction accuracy, the direct relationship between variables and model accuracy is not straightforward. In other words, a reduction in variables in a specific model may either decrease or increase the accuracy of model, depending on the relationships between variables and model types (Heung et al. 2014).

**DSM.** Although traditional maps are still a major source of information on the distribution of SOC, the development of DSM offers better ways to generate such information in the Czech Republic (Žížala et al. 2022). DSM provides essential tools to improve the understanding of the distribution of SOC for both forest and agricultural soils. It increases the efficiency of mapping process and provides a more detailed, accurate, and quantitative prediction of soil properties for different areas (Lorenzetti et al. 2015). Additionally, DSM has become a powerful tool for optimal decision-making in environmental and agricultural management by providing relevant soil information (McBratney et al. 2003). DSM integrates information from observed soil attributes with dependent environmental covariates obtained from terrain analysis, geospatial data sources, and remote sensing images using geographical information systems (GIS) and machine learning (ML) to generate grid-based maps of different soil types and properties and predict the spatial distribution of soil properties using a quantitative framework (McBratney et al. 2003; Mulder et al. 2011).

**Machine learning.** ML is the self-adaptive method where a fitted pattern can be used to set prediction targets for new data. Brungard et al. (2015) reported that covariates selected by soil scientists familiar with the study area did not yield the most accurate models compared to covariates automatically selected by ML algorithms. Additionally, Borůvka et al. (2022) showed that even large datasets used for modelling do not guarantee highly accurate prediction. Although the number of studies using ML algorithms have been increasing, only a few studies have compared different learners, and most studies are limited to the evaluation of a few common models such as random forest (Fathololoumi et al. 2020). QRF takes into account both landscape properties and the density, which is closer to the experience of a soil surveyor. QRF is based on the hypothesis that the clustering,

https://doi.org/10.17221/119/2023-SWR

optimizing the prediction of the mean, also optimizes the prediction of the other quantiles and the uncertainty. Although this has not been fully proven yet, Meinshausen (2006) showed that QRF clearly outperformed the quantile regression algorithms estimating each quantile separately in five different case studies (Vaysse & Lagacherie 2017).

## METHODOLOGY

An overall evaluation of the performance of RF, cubist, and QRF models for SOC mapping was conducted using R, version 3.5.1 (R Core Team 2018), to provide a framework for interpretation. The local uncertainty was assessed through a rigorous cross-validation approach. The evaluation consisted of comparing metrics of the performance, visual inspection, and interpretation of anomalies with geographic knowledge to figure out why the accuracy of the model for some locations in the landscape is not acceptable.

### Study area and soil sampling

The Liberec district (989 km$^2$) is located in the northern Czech Republic (Figure 1) with elevations ranging from 210 to 1 124 m above mean sea level and is covered by 47.2% agricultural land, 42.4% forest area, and less than 6% mix of agriculture and forests (Miko & Hošek 2009), as illustrated in Figure 3C. On the other hand, the Domažlice district (1 123 km$^2$) is located in the western Czech Republic with elevations ranging from 383 to 1 042 m and is covered by 53.0% agricultural land, 38.2% forest land, and less than 6% mix of agriculture and forests (Miko & Hošek 2009), as illustrated in Figure 4C.

The soil was classified according to the Czech Taxonomic Soil Classification System and WRB system (Němeček et al. 2011; IUSS Working Group WRB 2015). Eight classes of mineral parent material including sedimentary rocks, acid granites and similar rocks, basalts, loess-like sediments, micaceous schists and phyllites, polygenetic loams, gneisses, alluvial (fluvial) and six major reference soil groups including Cambisols, Podzols, Gleysols, Stagnosols, Luvisols, and Fluvisols were identified in the Liberec district. Eight classes of mineral parent material including sedimentary rocks, acid granites and similar rocks, other mafic rocks, loess-like sediments, micaceous schists and phyllites, polygenetic loams, gneisses, alluvial (fluvial) sediments as well as five major reference soil groups including Cambisols, Gleysols, Stagnosols, Luvisols, and Fluvisols were observed in the Domažlice district (Němeček et al. 2011; IUSS Working Group WRB 2015).

In this study, 71 samples for the Liberec and 67 samples for the Domažlice districts were randomly collected in 2004 (Figure 2 and Table 1).

The sampling depth was 0–30 cm because it represents the plow depth and SOC estimation in this depth is an important factor in farm management. In each location, soil was sampled to a depth of 30 cm using a steel soil auger after removing plant debris such as grass and twigs. In forest, the forest floor was also removed for consistency of samples across different land covers. The collected soil samples were stored in plastic bags and transferred to the laboratory for analysis. To measure SOC, soil samples were air-dried, grinded and sieved using a sieve with mesh size < 0.25 mm, and the SOC was determined through the oxidimetric modified Tyurin method (Pospíšil 1964).

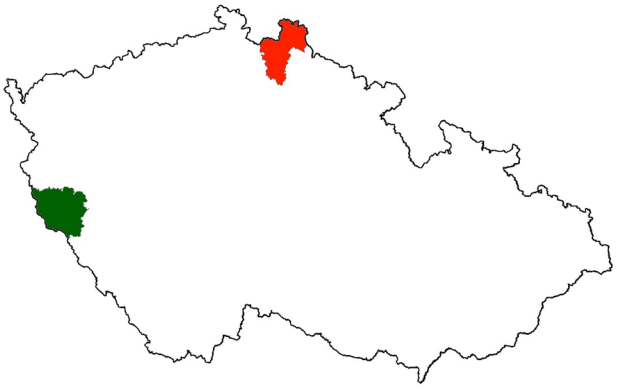

Figure 1. Location of the Liberec (red) and Domažlice (green) districts in the Czech Republic

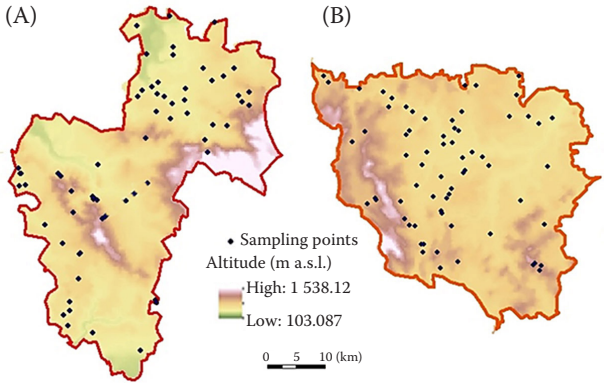


Figure 2. Elevation maps and distribution of sampling locations in the Liberec (A) and Domažlice (B) districts

https://doi.org/10.17221/119/2023-SWR

Table 1. Number of samples collected in different land-uses

| District | Agricultural land | Forest area | Mix of agricultural and forest lands | Total |
|---|---|---|---|---|
| Liberec | 35 | 22 | 14 | 71 |
| Domažlice | 35 | 20 | 12 | 67 |

**Legacy data and auxiliary environmental covariates**

Environmental covariates are essential in the DSM process and can be obtained from a combination of remotely sensed data, digital elevation model (DEM), or other geospatial sources (Lagacherie et al. 2006). Soil survey and soil mapping have a long tradition in the Czech Republic. Various large-scale point or polygon legacy soil data and maps are available in the country (Kozák et al. 1996; Němeček 2000) and Europe (Panagos et al. 2014). In this research, a DEM with a 100 m spatial resolution was obtained from the U.S. Geological Survey database (USGS 2021). A suite of 15 topographic variables was computed using SAGA GIS 7.2.0 (Conrad et al. 2015). In addition, normalized difference vegetation index (NDVI) (Landsat TM image (USGS 2021)), soil, geological (Kozák et al. 1996), and CORINE Land Cover maps (EEA 2018) were used as the predictors. The CORINE database contains four main categories including forest, arable land, pasture, and industrial areas (Figures 3 and 4). Table 2 presents a summary of the total 16 environmental variables used in this study. Borůvka et al. (2022) reported that the importance of environmental variables in the models for SOC stock prediction varies in different regions and altitudes.

Vegetation indices are helpful in modelling SOC because the vegetation is the ultimate source of SOC. NDVI is a common unitless remote sensing index that uses the ratio between visible and near-infrared reflectance of vegetation cover. Additionally, it can estimate the green density of the area (Weier & Herring 2000). NDVI is calculated as:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

where:

NIR – the amount of image reflection in the near-infrared band;

RED – is the amount of image reflection in the red band.

NDVI was calculated from a Landsat TM image with < 10% cloud cover. The image was taken in 1992 under clear weather conditions on the 9$^{th}$ of August for the Domažlice district and the 19$^{th}$ of September for the Liberec district. Two spectral bands were selected from Landsat Legacy TM including Band three, containing red reflectance, and Band four with infrared reflectance.

**Basic statistical analyses**

Statistical differences in mean values were computed by one-way analysis of variance (ANOVA) method using SPSS (SPSS 2001) and R, version 3.5.1 (R Core Team 2018). One-way ANOVA was conducted to evaluate the effects of landform types (slope) on soil properties (Duncan's test at the 5% level of significance). The SPSS analysis was also carried out to determine the correlation matrix between variables used in this study. In addition, multiple and linear regression coefficients were calculated to determine the relationships between auxiliary variables and SOC using R and SPSS.

**Regression models**

The models used in this study include a tree-based methods called RF, cubist, and QRF (Pahlavan-Rad et al. 2020). Many studies have demonstrated that RF has superior performance compared to other models (Brungard et al. 2015; Pahlavan-Rad et al. 2018; Zeraatpisheh et al. 2019). Indeed, RF is a modified and extended model of the regression tree model (as a basic idea) and it constructs a forest of low-correlation regression trees (Peters et al. 2007). However, the original implementation of RF was unable to produce spatial estimates of uncertainty. Therefore, QRF was introduced as an alternative to the RF learner, allowing users to leverage the model predictions from each tree of the RF to generate uncertainty estimates. On the other hand, cubist is a modification and extension of the basic classification tree idea (Quinlan 1993).

Although many studies have proved the high accuracy of the RF model, only a few works have shown that the performance of the RF model is not perfectly acceptable (Pouladi et al. 2019). In addition, one of the limitations of the RF model is that the soil properties may be overestimated (Pahlavan- Rad & Akbari Moghaddam 2018).

**Random forest (RF).** RF algorithm, proposed by Breiman (2001), is an ensemble learner which consists of many individual decision trees which are built

https://doi.org/10.17221/119/2023-SWR

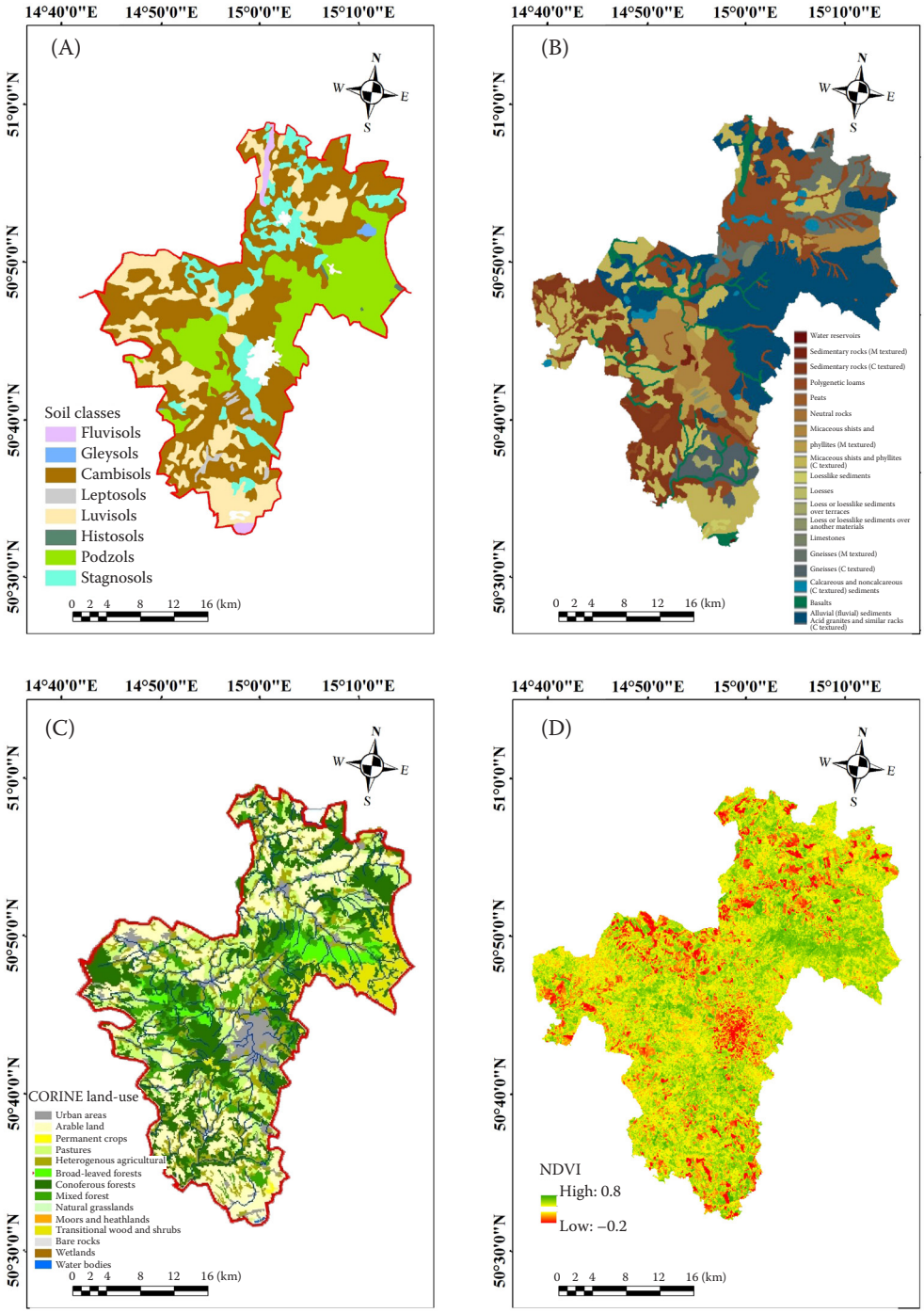


Figure 3. Soil map (A), geology map (B), land cover (C), and normalized difference vegetation index (NDVI) (D) for the Liberec district

from a bootstrap sample taken from the population of all samples, *ntree*. Additionally, the node-splitting rules are generated by randomly selecting a predictor from a subset of predictors based on *mtry*, which is the main tuning parameter for RF. *Mtry* and *ntree* were identified as those returning the lowest out of bag (OOB) error by iterating *mtry* values from one to the total number of important variables and *ntree* values were chosen 1 000. The results of individual models are aggregated into an ensemble using an averaging function when predicting continuous response variables (i.e., SOC). The ensemble modelling approach

https://doi.org/10.17221/119/2023-SWR

is designed to mitigate the impacts of model overfitting. The RF model was implemented using the Caret package (Kuhn 2012; Brewer et al. 2015) in the R statistical software (R Core Team 2018).

**Quantile random forest (QRF).** QRF is an extension of the RF learner, which allows users to leverage the model predictions from each tree of the RF to generate uncertainty estimates. Meinshausen (2006) reported that QRF not only provides information about the conditional mean, but it also provides information about the conditional distribution of the target variable. In addition, only the mean of the

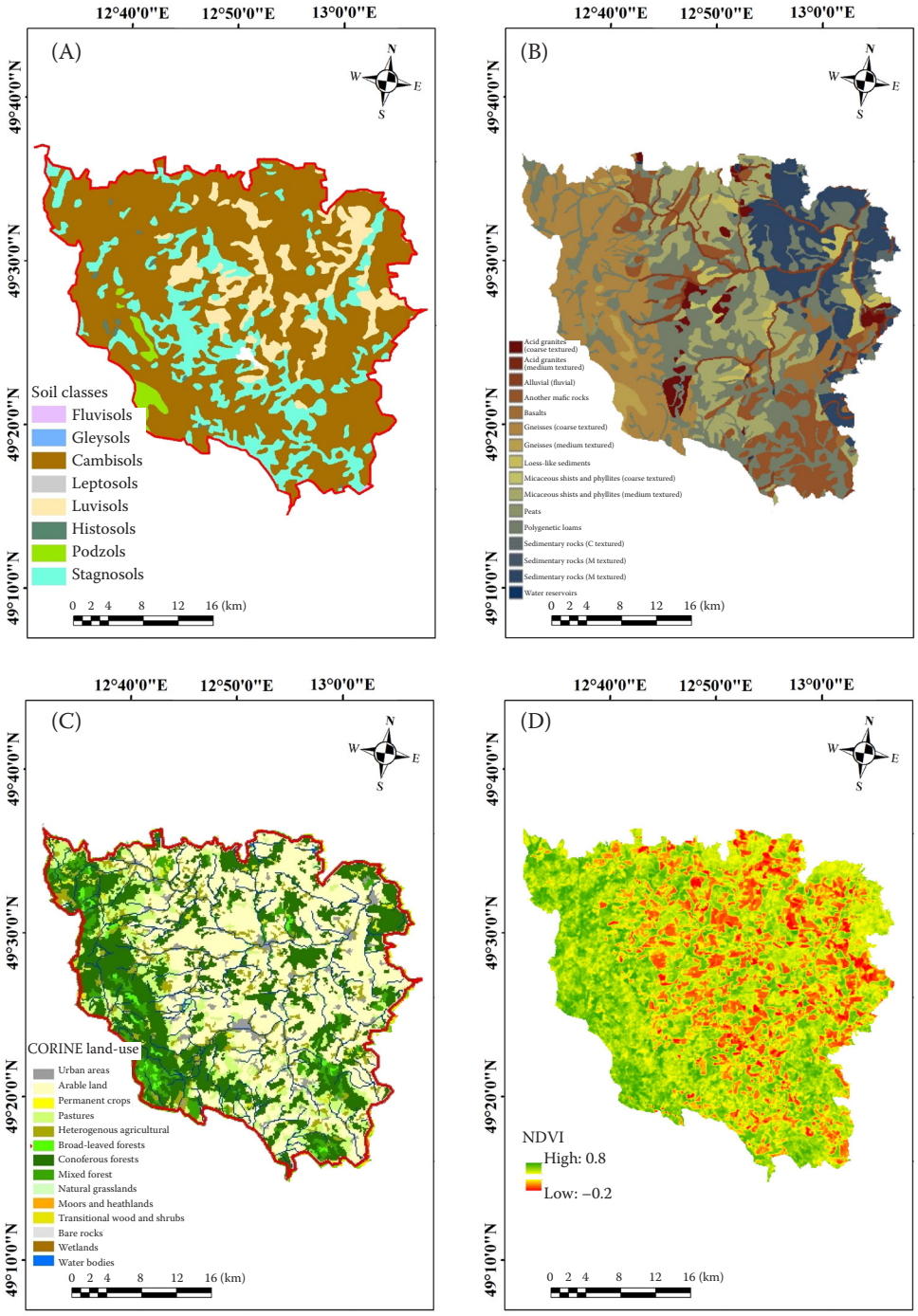


Figure 4. Soil map (A), geology map (B), land cover (C), and normalized difference vegetation index (NDVI) (D) for the Domažlice district

https://doi.org/10.17221/119/2023-SWR

Table 2. Soil environmental covariates mostly derived from digital elevation model (DEM) (McBratney et al. 2003)

| Soil-environmental covariates | Code | Significance related to soil development and properties |
|---|---|---|
| Elevation | Elev | climate, vegetation, energy potential |
| Slope | S | surface and subsurface flows, flow speed and erosion rate, precipitation, vegetation, geomorphology, soil water content, land-use capacity |
| Profile curvature | PC | profile curvature is the rate of change of slope in a downslope direction; it characterizes changes in flow acceleration that may differentiate erosion and deposition zones in landscapes |
| Plan curvature | Plan. Cur | convergent/divergent flows, soil water content, soil characteristics, flow acceleration, erosion rate/deposition, geomorphology |
| Length slope factor | LSF | surface flow volume |
| Topographic wetness index | TWI | a measure of the topographic control on soil wetness |
| Valley depth | Va. Dep | valley depth specifies soil characteristics, influencing composition and fertility, crucial for effective land management |
| Relative slope position | RSP | it is a measure of the percentage distance a location is from slope bottom to nearest ridge top, influencing drainage, erosion, and microenvironments |
| Convergence index | CI | it is calculated based on the aspect that shows the structure of the relief and flow convergence affecting water movement |
| Vertical distance to channel networks | VDCN | a grid provides information about the channel network, influencing drainage patterns and sediment transport |
| Channel network base level | CNBL | this grid output contains the interpolated channel network of base level elevations, defining landscape lowering and drainage efficiency |
| Total catchment area | Cat. Area | expected runoff volume that determines water inflow and sediment transport |
| Normalized difference vegetation index of Landsat-4 | NDVI | it reflects vegetation health and biomass |
| Geology map | Geology | polygon map of soil parent material |
| Soil map | Soil | polygon map of soil classes |
| Land-use map | Land-use | polygon map of CORINE land cover categories that shows vegetation and human activities impacting soil |

observations within the terminal node is used in RF whereas QRF keeps all predicted values for each terminal node. Accordingly, QRF retains the residual distribution at each terminal node, which is used to estimate the prediction interval width. The QRF model was implemented using the Quantreg Forest package (Vaysse & Lagacherie 2017) used by R Studio 3.5.0 software.

**Cubist.** Cubist is an extension of Quinlan's M5 model tree (Quinlan 1993) and was implemented using the cubist package in R (Kuhn et al. 2013; R Core Team 2018). Although cubist is comparable to ordinary regression trees, its leaves are in the form of a linear regression equation (Taghizadeh-Mehrjardi et al. 2016). Considering the hybridization of a tree-based model with linear models, cubist can characterize both linear and non-linear relationships. It is also worth mentioning that many researchers have been using the cubist model in different soil prediction and mapping techniques (for example Henderson et al. 2005; Minasny et al. 2008). The cubist method's principal achievement is to use multiple training committees and boosting to make the weights more balanced. The outstanding usage of cubist is to analyse enormous databases that include a great number of records and numeric or nominal fields. Cubist models also compute variable importance to model accuracy as a variable's relative contribution.

## Uncertainty

Uncertainties in SOC stock assessments are critical in determining the significance of the results. No prediction is free from errors, as every model is a simplified representation of reality. The prediction error can be tracked down to uncertainty introduced in a model either as a result of input uncertainty or during incomplete construction of a model. Therefore, the modelling process is very dependent on training

https://doi.org/10.17221/119/2023-SWR

data, not only because of its uncertainties, but also because QRF estimates the cumulative distribution function (CDF) by using an empirical CDF. Therefore, it quantifies the complete error given a certain input vector as it includes a conditional variance estimate for Y by using the information within the leaves. Hence, Meinshausen's et al. (2016) technique can be used for making prediction intervals and not for confidence intervals because the empirical CDF provides no information on the uncertainty of the fit of the RF model itself since the data needs to be a representative sample of the underlying populations (James et al. 2014).

**Model evaluation**

Spatial models were validated by leave-group-out cross-validation (LGOCV) as well as by independent validation. The latter was performed by randomly splitting the sample set (70% calibration and 30% validation). Each model was fitted using the train data and the test data was used for validation. Differences between observed and predicted values were summarized as the root-mean-squared error of prediction (RMSEP) and the bias of the estimation.

Model evaluation is an essential factor for accurately predicting SOC (Mosleh et al. 2016). K-fold cross-validation is usually used to evaluate model performance. In this research, the training dataset was randomly partitioned into 10 folds ($k$ = 10), so 10-fold cross-validation was used. The model was trained using $k$ = 9 folds, tested with the one remaining fold, and accuracy metrics were calculated based on the test fold. The process of training and testing was repeated 10 times so each individual fold was selected as the test set once. The model performance was evaluated based on the average accuracy metrics of all folds including mean absolute prediction error (MAE), root mean square error (RMSE), and index of determination ($R^2$). The metrics were used to evaluate the prediction error rates and model efficiency as well as to investigate the correspondence between predicted and measured data.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \quad (2)$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \quad (3)$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \overline{y}\right)^2} \quad (5)$$

where:

$y_i$ – the measured value at $i$-th location;

$\hat{y}_i$ – the predicted value of $y$;

$\overline{y}$ – the mean value of $y$;

$N$ – the number of units (locations).

Although $R^2$ is a valid statistic to evaluate the prediction accuracy of a model, a high $R^2$ may not lead to accurate predictions. This is because the model could systematically and considerably over- and/or under-estimate the data at different points along the regression line. As a result, evaluation of the models using other performance statistics appears to be necessary to provide complement information on prediction accuracy. A lower gained value is adequate and evaluated best for the selection of a model using the RMSE and MAE validation criteria evaluation methods. In the current study, the Li et al. (2016) criterion was applied. They proposed a classification criterion for $R^2$: unacceptable prediction ($R^2 < 0.50$), acceptable prediction ($0.50 \leq R^2 < 0.75$), and good prediction ($R^2 \geq 0.75$).

## RESULTS AND DISCUSSION

**Summary statistics and correlation analysis**

Statistics summary of SOC is presented in Table 3. The average SOC was 2.83% and 2.83% in the Liberec and Domažlice districts, respectively (Nozari & Borůvka 2023). The results of the correlation analy-

Table 3. Statistics summary of soil organic carbon (SOC) for both districts (%)

| District | No. of observations | Minimum | Maximum | Mean | Median | SD |
|---|---|---|---|---|---|---|
| Liberec | 71 | 0.42 | 11.33 | 2.83 | 1.86 | 2.50 |
| Domažlice | 67 | 0.00 | 9.33 | 2.83 | 1.49 | 2.39 |

SD – standard deviation

https://doi.org/10.17221/119/2023-SWR

sis between environmental covariates and SOC are also presented in Tables 4 and 5. SOC was positively correlated with elevation. Although the correlation between most of the variables and SOC was not high, the RF model can identify nonlinear relationships between variables. Nevertheless, one of the significant parameters to explain SOC content variation is elevation, particularly in areas outside of flat sub-humid climates (Tziachris et al. 2019). The characteristics showing good correlation with SOC indicate potential candidates for strong predictors in SOC models, as analysed in section Variable importance.

**Model validation**

Table 6 presents values for $R^2$, MAE, and RMSE in the Liberec and Domažlice districts using RF, QRF, and cubist models. $R^2$ values ranged between 0.40 and 0.68, MAE ranged between 0.98 and 1.49, and RMSE ranged between 1.32 and 2.21. It is generally believed that $R^2$ values greater than 0.4 indicate the effectiveness of the model in predicting soil properties (Prasad et al. 2006; Moore et al. 2013). Although $R^2$, MAE, and RMSE values for all three models used in this study were similar, RF consistently showed greater accuracy metrics (greater $R^2$ but smaller RMSE and MAE values) compared to cubist and QRF for both districts. Therefore, based on these accuracy metrics RF was considered the most accurate algorithm among the three models used in this research. This finding supports the findings of other studies which also found that RF is suitable for soil spatial modelling due to its high accuracy (Ellili et al. 2019; Lamichhane et al. 2019). It should be noted that these indicators may not be suitable for prediction accuracy of the local uncertainty.

**Variable importance**

Mosleh et al. (2016) reported that the parameters derived from the DEM in low-relief areas are appropriate environmental factors to model soil properties. In this research, the relative importance of the predictor variables in the SOC modelling was evaluated using the Varlmp function in R (R Core Team 2018). It is believed that the climate, temperature, and disaster conditions (e.g. large-scale geological or meteorological events such as flooding, runoff, erosion, drought, and dust storms) are similar in both districts. In the Liberec district, the most effective variables in SOC prediction using RF and cubist models were land cover (vegetation), elevation, valley depth, and slope as illustrated in Figure 5. Similarly, Ellili et al. (2019) found that slope and elevation are the most important covariate variables for predicting SOC. These results indicate that the land cover is essential in identifying the SOC distribution. Coniferous forest, broad-leaved forest, and mixed forest containing a great amount of SOC in both districts show the importance of forest management. In addition, slope and valley depth, which are related to the topography and hydrology of the region, have effects on water distribution and runoff transport, affecting the erosion and deposition that changes the spatial variation of SOC in this mountainous area, as well as the soil organic matter (SOM) decomposition and accumulation processes. Generally, the steeper and longer a slope is, the faster water runs off from it, increasing the potential of erosion. Therefore, the influences of slope and valley depth were also highlighted as significant auxiliary variables in predicting SOC in the Liberec region. In the Domažlice district, the most important variables in SOC prediction using RF and cubist models were land cover and elevation as illustrated in Figure 5. It also confirms the significance of vegetation and topographic parameters similar to the Liberec district.

**Spatial prediction of SOC**

The spatial distribution of SOC content for the topsoil in the Liberec and Domažlice districts using RF, cubist, and QRF models is illustrated in Figures 6, 7, and 8, respectively.

Generally, all three models used in this study were similar in terms of spatial patterns of SOC content. The biggest SOC content was found in high elevations covered by forests, while the lowest SOC content was observed in the areas where croplands have replaced the plantation and indigenous forests, which is consistent with observations by other researchers (Tesfahunegn et al. 2011; Winowiecki et al. 2016). Therefore, a reduction in SOC stocks could be due to the biomass removal after harvesting, erosive processes, and frequent tillage that breaks up the soil aggregates, alters aeration, and accelerates the microbial decomposition and oxidation of SOM to $CO_2$. It was also found that increasing elevation increased the average SOC concentrations, confirming that SOC responds to climatic variables such as temperature that decreases as elevation increases. Increased SOC may also be due to the recent changes in land-use. For example, agricultural lands at higher elevations are more likely to have been recently changed to another type of land (grassland).

https://doi.org/10.17221/119/2023-SWR

Table 4. Correlation matrix (Pearson) for the Liberec district

| Variables | SOC | Elev | S | Aspect | Sin (aspect) | Cos (aspect) | Plan. cur | PC | CI | Cat. area | TWI | LSF | CNBL | VDCN | RSP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SOC | 1 | | | | | | | | | | | | | | |
| Elev | 0.380** | 1 | | | | | | | | | | | | | |
| S | 0.525** | 0.667** | 1 | | | | | | | | | | | | |
| Aspect | –0.115 | –0.127 | –0.140 | 1 | | | | | | | | | | | |
| Sin (aspect) | 0.129 | 0.213 | 0.185 | –0.836 | 1 | | | | | | | | | | |
| Cos (aspect) | –0.085 | 0.139 | 0.158 | –0.137 | 0.199 | 1 | | | | | | | | | |
| Plan. cur | 0.049 | 0.491** | 0.352 | –0.049 | 0.061 | 0.116 | 1 | | | | | | | | |
| PC | –0.179 | 0.322 | 0.050 | –0.037 | 0.013 | –0.048 | 0.746** | 1 | | | | | | | |
| CI | 0.079 | 0.314 | 0.156 | –0.077 | 0.141 | 0.103 | 0.708** | 0.591 | 1 | | | | | | |
| Cat. area | 0.046 | –0.254 | 0.104 | 0.071 | –0.146 | –0.185 | 0.031 | –0.047 | –0.018 | 1 | | | | | |
| TWI | –0.145 | –0.598 | –0.530 | 0.134 | –0.239 | –0.228 | –0.521 | –0.402 | –0.490 | 0.421 | 1 | | | | |
| LSF | 0.508** | 0.440 | 0.917** | –0.053 | 0.069 | 0.065 | 0.242 | –0.056 | 0.085 | 0.310 | –0.299 | 1 | | | |
| CNBL | 0.100 | 0.568** | 0.249 | 0.021 | –0.017 | 0.100 | 0.014 | –0.109 | –0.070 | –0.248 | –0.118 | 0.190 | 1 | | |
| VDCN | –0.206 | –0.487 | –0.307 | –0.124 | –0.012 | –0.020 | –0.498 | –0.526 | –0.351 | 0.171 | 0.513** | –0.129 | 0.125 | 1 | |
| RSP | 0.387** | 0.842** | 0.574** | –0.077 | 0.180 | 0.077 | 0.568** | 0.490** | 0.394 | –0.203 | –0.667 | 0.341 | 0.121 | –0.806 | 1 |

SOC – soil organic carbon; Elev – elevation; S – slope; Plan. cur – plan curvature; PC – profile curvature; CI – convergence index; Cat. area – total catchment area; TWI – topographic wetness index; LSF – length slope factor; CNBL – channel network base level; VDCN – vertical distance to channel network; RSP – relative slope position; *,**,***correlation is significant at $P < 0.05$, 0.01, 0.001, respectively

Soil and Water Research, 19, 2024 (1): 32–49

https://doi.org/10.17221/119/2023-SWR

Table 5. Correlation matrix (Pearson) for the Domažlice district

| Variables | SOC | Elev | S | Aspect | Sin (aspect) | Cos (aspect) | Plan. cur | PC | CI | Cat. area | TWI | LSF | CNBL | VDCN | RSP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SOC | 1 | | | | | | | | | | | | | | |
| Elev | 0.658** | 1 | | | | | | | | | | | | | |
| S | 0.444** | 0.721 | 1 | | | | | | | | | | | | |
| Aspect | 0.030 | –0.087 | –0.156 | 1 | | | | | | | | | | | |
| Sin (aspect) | –0.092 | 0.034 | 0.055 | –0.833 | 1 | | | | | | | | | | |
| Cos (aspect) | –0.096 | –0.012 | 0.097 | 0.002 | –0.057 | 1 | | | | | | | | | |
| Plan. cur | 0.253 | 0.449 | 0.235 | –0.339 | 0.377 | –0.063 | 1 | | | | | | | | |
| PC | –0.064 | 0.094 | 0.016 | –0.194 | 0.184 | 0.003 | 0.547 | 1 | | | | | | | |
| CI | 0.167 | 0.333 | 0.153 | –0.241 | 0.238 | –0.160 | 0.709 | 0.374 | 1 | | | | | | |
| Cat. area | –0.134 | –0.265 | –0.156 | –0.094 | 0.060 | 0.001 | –0.075 | –0.055 | –0.170 | 1 | | | | | |
| TWI | –0.347 | –0.596 | –0.685 | 0.103 | –0.124 | –0.007 | –0.457 | –0.240 | –0.508 | 0.569 | 1 | | | | |
| LSF | 0.398** | 0.644 | 0.962 | –0.109 | –0.014 | 0.047 | 0.079 | –0.097 | 0.040 | –0.091 | –0.532 | 1 | | | |
| CNBL | 0.449** | 0.718 | 0.416 | 0.076 | –0.112 | 0.073 | 0.031 | –0.183 | 0.053 | –0.309 | –0.324 | 0.410 | 1 | | |
| VDCN | –0.060 | –0.099 | –0.103 | 0.082 | –0.047 | 0.037 | –0.239 | –0.397 | –0.114 | 0.031 | 0.093 | –0.072 | 0.181 | 1 | |
| RSP | 0.556** | 0.805** | 0.652** | –0.199 | 0.121 | –0.003 | 0.521 | 0.323 | 0.398 | –0.238 | –0.588** | 0.549** | 0.301 | –0.528 | 1 |

SOC – soil organic carbon; Elev – elevation; S – slope; Plan. Cur – plan curvature; PC – profile curvature; CI – convergence index; Cat. Area – total catchment area; TWI – topographic wetness index; LSF – length slope factor; CNBL – channel network base level; VDCN – vertical distance to channel network; RSP – relative slope position; *,**,***correlation is significant at $P < 0.05, 0.01, 0.001$, respectively

https://doi.org/10.17221/119/2023-SWR

Table 6. Assessment results for random forest (RF), cubist, and quantile random forest (QRF) models for soil organic carbon (SOC) prediction

| Model | Liberec district | | | Domažlice district | | |
|---|---|---|---|---|---|---|
| | $R^2$ | MAE | RMSE | $R^2$ | MAE | RMSE |
| RF | 0.58 | 1.40 | 1.98 | 0.68 | 0.98 | 1.32 |
| QRF | 0.48 | 1.28 | 1.98 | 0.49 | 1.18 | 1.74 |
| Cubist | 0.40 | 1.49 | 2.21 | 0.64 | 1.08 | 1.57 |

$R^2$ – coefficient of determination; MAE – mean absolute error; RMSE – root mean square error

Results also showed greater SOC accumulation extending from northeast to west areas of the Liberec region, and west to southeast parts of the Domažlice region. Generally, elevation changes affect the soil physicochemical attributes which are the main factors for predicting the SOC content variation. As can be seen in Figures 6, 7, and 8, SOC content has changed sharply based on QRF, while RF shows a more continuous distribution of SOC content for both regions. In addition, QRF maps showed different SOC distribution from RF and cubist maps in the Liberec district. SOC content in the western part of the study areas predicted by QRF maps is much lower than that predicted by cubist. Although RF maps show a rather even distribution of SOC content in agricultural and forestry areas, its results greatly differed from the cubist and QRF maps. RF map indicated that SOC content of the elevation ridge in the eastern parts of the Domažlice district is much lower than that predicted by cubist. Moreover, the RF map predicted a higher amount of SOC in northern parts, southern parts, and around the centre of the Liberec district than cubist and QRF maps.

Topography maps (Figure 2) show that the Liberec district is mostly mountainous and sloping, which results in a greater accumulation of SOC. This can

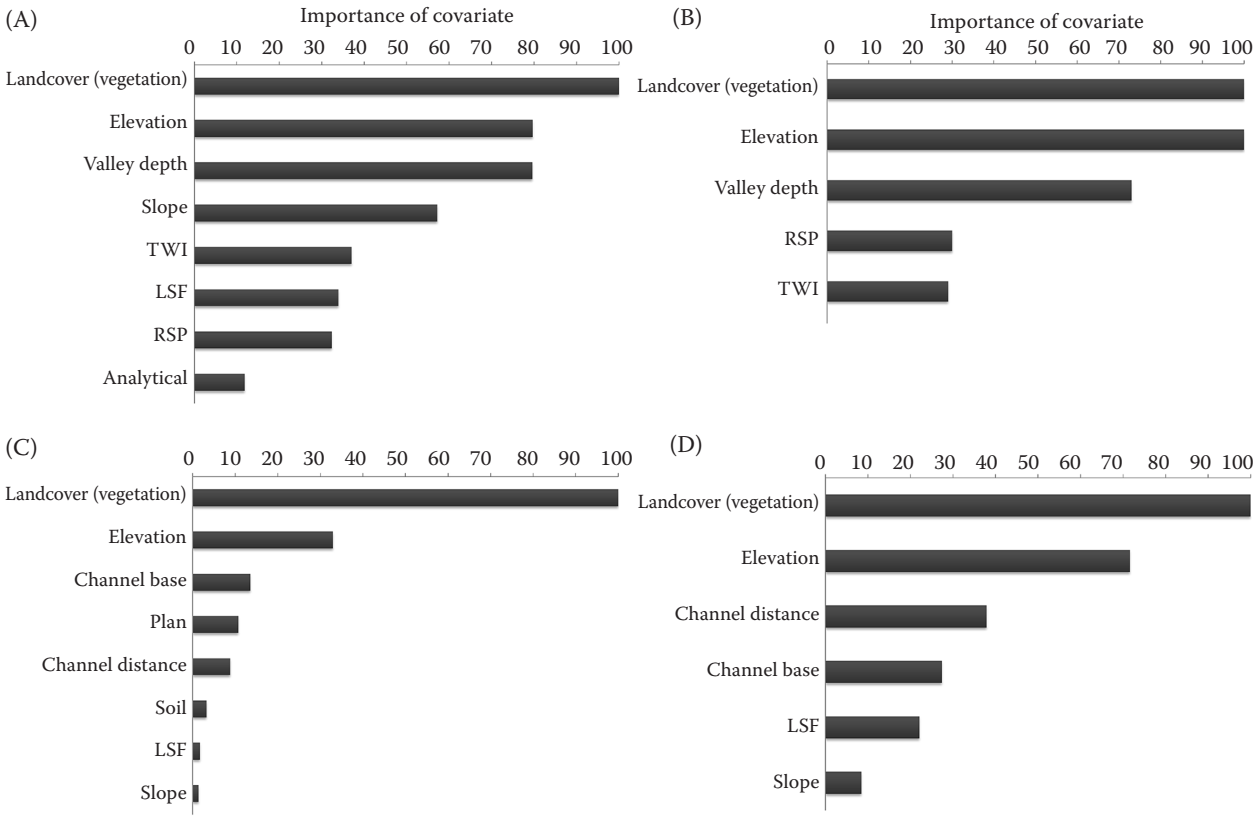


Figure 5. Relative variable importance (%) for soil organic carbon (SOC) spatial prediction by random forest (RF) in Liberec (A), cubist in Liberec (B), RF in Domažlice (C), and cubist in Domažlice (D)
TWI – topographic wetness index; LSP – length slope factor; RSP – relative slope position

https://doi.org/10.17221/119/2023-SWR

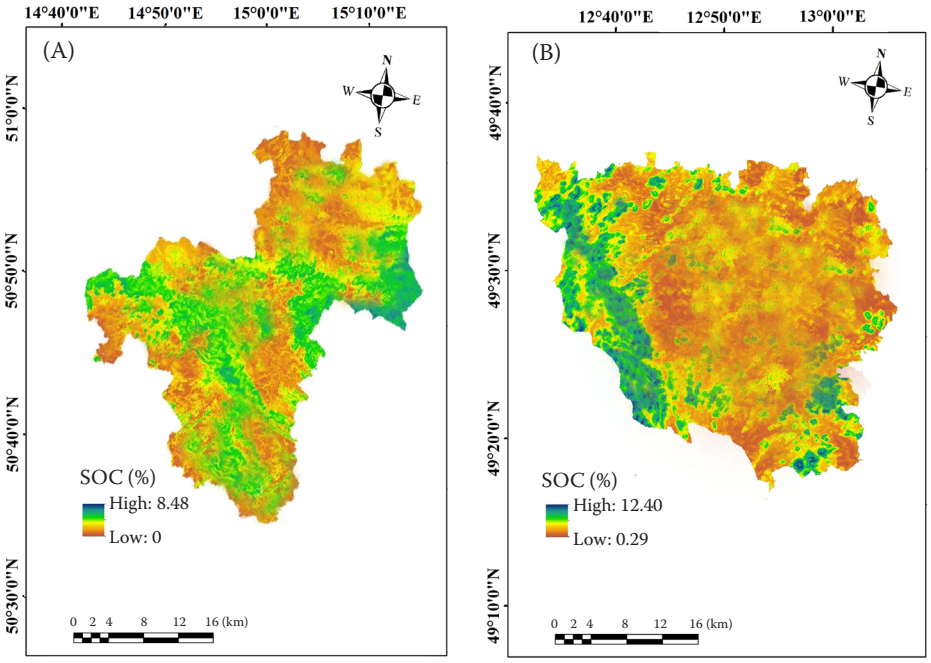


Figure 6. Soil organic carbon (SOC) distribution maps using random forest (RF) model in the Liberec (A) and Domažlice (B) districts

be due to the combined effects of soil acidification through reduced decomposition in higher elevations and poor water drainage on lower slopes. Similarly, Zhu et al. (2018) reported that the SOC content is more aggregated and less decomposed in soils with greater slope and poor drainage. In addition, topographic variables such as the depth of the valley have affected water distribution, runoff velocity, and

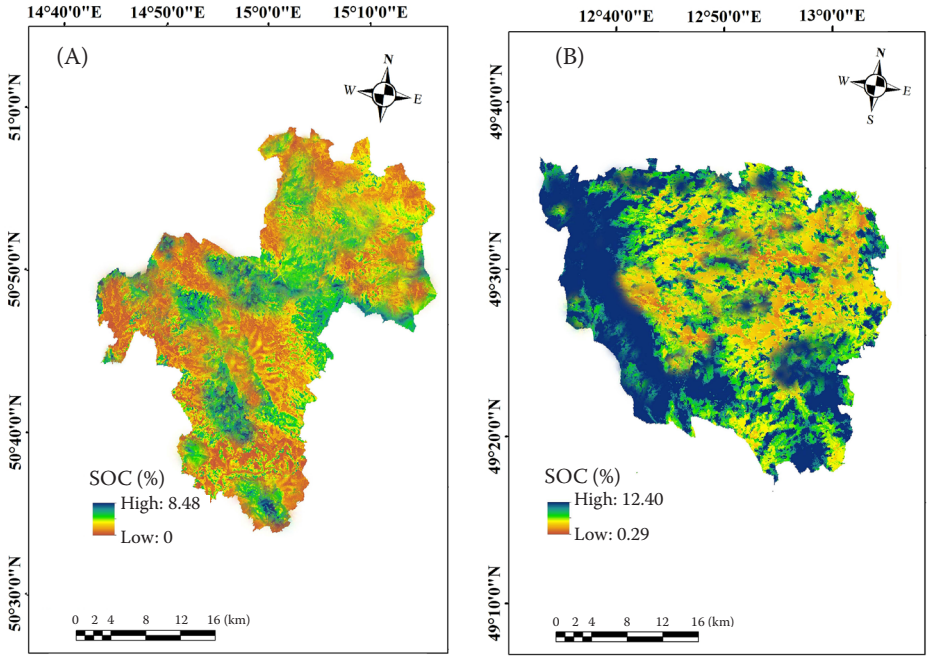


Figure 7. Soil organic carbon (SOC) distribution maps using cubist model in the Liberec (A) and Domažlice (B) districts

https://doi.org/10.17221/119/2023-SWR

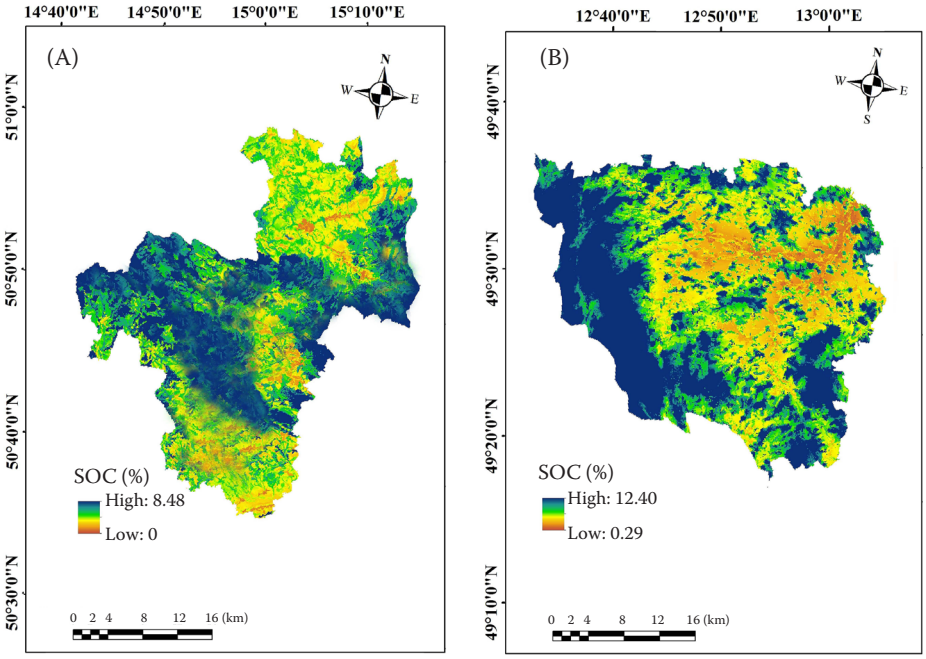


Figure 8. Soil organic carbon (SOC) distribution maps using quantile random forest (QRF) model in the Liberec (A) and Domažlice (B) districts

sediment erosion, and so the spatial variation of SOC in both regions has increased. The reduction of SOC is more pronounced in agricultural and residential areas than in areas where human manipulation in nature is limited. Total SOC content in both districts is high due to the high humidity, forest vegetation such as coniferous forest, and a great amount of rainfall resulting in denser vegetation. As a result, this study

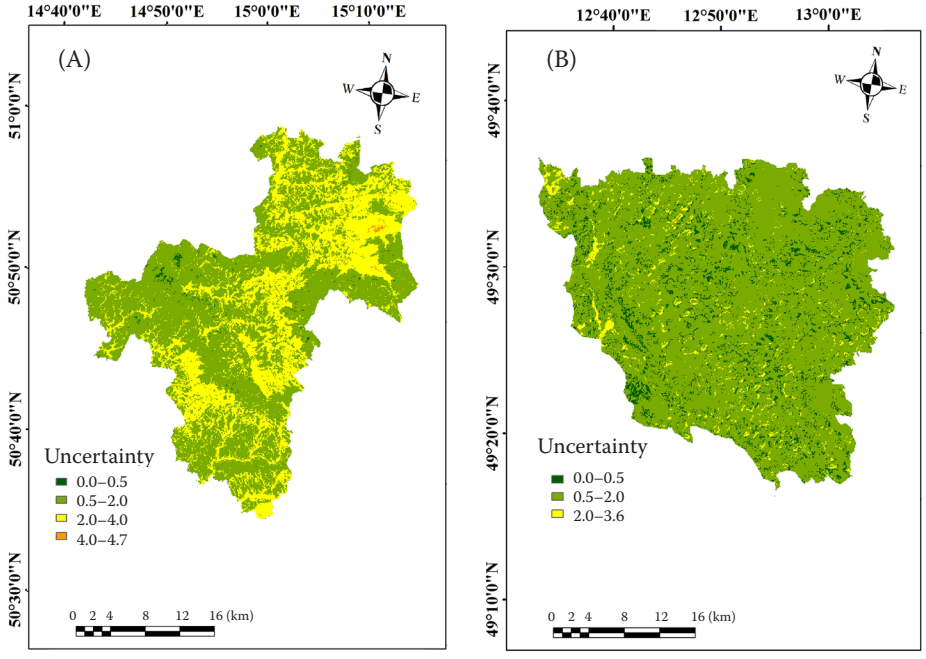


Figure 9. Soil organic carbon (SOC) uncertainty maps using random forest (RF) model in the Liberec (A) and Domažlice (B) districts

https://doi.org/10.17221/119/2023-SWR

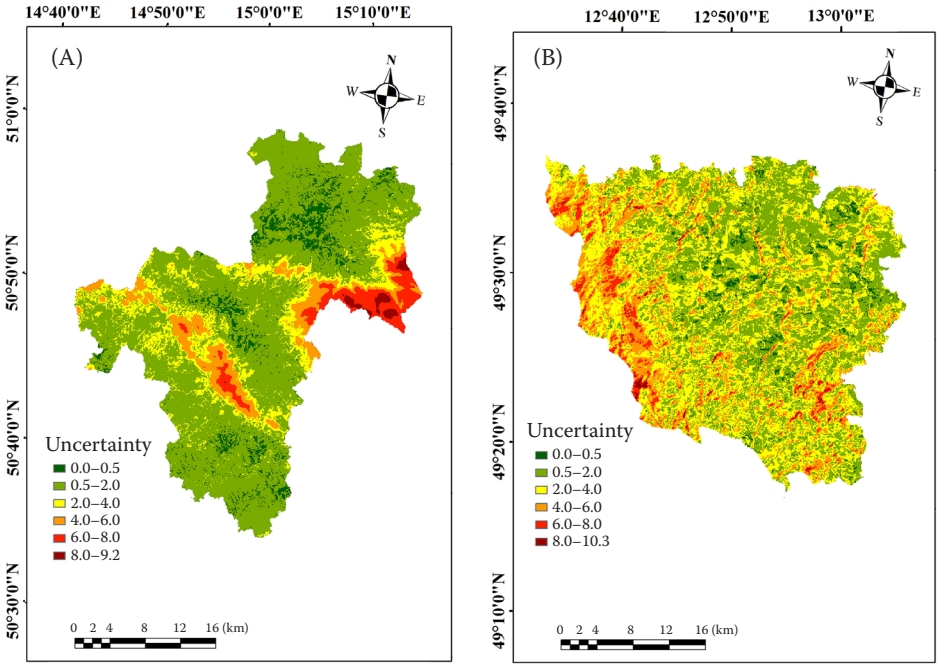


Figure 10. Soil organic carbon (SOC) uncertainty maps using cubist model in the Liberec (A) and Domažlice (B) districts

confirms the importance of terrain-based covariates and vegetation on SOC content variability in sub-humid areas. Also, there is a great variety of land-use and agricultural practices that may generate contrasting organic matter levels.

## SOC prediction uncertainty

DSM requires field observations, empirical prediction models, and a variety of environmental covariates to model spatially explicit predictions of soil properties. Therefore, the predictions are

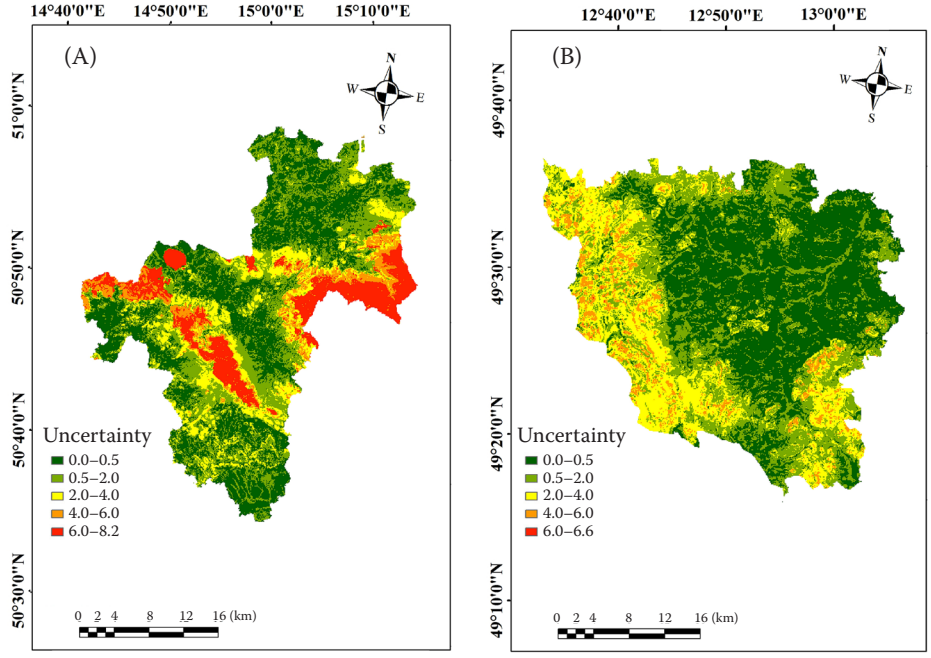


Figure 11. Soil organic carbon (SOC) uncertainty maps using quantile random forest (QRF) model in the Liberec (A) and Domažlice (B) districts

https://doi.org/10.17221/119/2023-SWR

always related to uncertainties brought by these three sources. SOC prediction uncertainty maps using RF, cubist, and QRF models for the Liberec and Domažlice districts are illustrated in Figures 9–11. The greatest uncertainty of SOC was in the coniferous forests in both districts (compared to landcover maps in Figures 3C and 4C). This is likely because there were relatively few SOC samples (15 samples in Liberec and 17 samples in Domažlice districts) and the variability of the covariates increased in these areas. Therefore, it is recommended to take more samples from these areas to ensure lower SOC prediction uncertainty. Interestingly, although each model had similar RMSE, MAE, and $R^2$, there are differences in uncertainty patterns in each model prediction for both districts. Particularly, differences between RF and QRF, using the same basic algorithm, were surprising. This can be because they differ in how the terminal nodes are dealt with. QRF appears to produce a lower uncertainty at lowlands compared to RF, however, both QRF and cubist show larger values in the mountainous regions where the models are most likely extrapolating. While the QRF model had low RMSE and MSE values similar to RF, the uncertainty distribution is much more uniform for RF.

Similar pattern is repeated in the Domažlice district as well. The higher uncertainty in predicted SOC was observed in mountains and forests compared to the cropland and pastures which is visible in western areas.

## CONCLUSION

This study assessed the spatial distribution of SOC in the Liberec and Domažlice districts in the Czech Republic. From the results of this study, the following conclusions can be drawn:

(1) Although the studied models including RF, QRF, and cubist did not have substantially excellent performance (not achieving high $R^2$ values), RF model consistently showed the best performance among all three models in both districts.

(2) Based on the RF model results, land cover (vegetation) and elevation were the most important environmental covariates for the prediction of SOC in both districts.

(3) The highest SOC content was predicted in the highest elevation in the forest-dominated areas (northeastern to western parts of the Liberec region and western to southeastern parts of the Domažlice region) while the lowest SOC was found in the lowest elevations in the cropland-dominated areas.

(4) The greatest uncertainty of SOC was observed in the coniferous forests in both districts, most likely because there were relatively few SOC samples and the variability of the covariates increased in these areas.

(5) Overall, RF can use many terrain covariates which have a strong spatial association with SOC and is considered the most accurate predictive model for both districts because it showed better performance (greater $R^2$ but smaller RMSE and MAE values, more uniform uncertainty distribution without very high uncertainty values) compared to cubist and QRF for both districts.

(6) Finally, to improve the prediction accuracy of SOC distribution, more observations and stratified random sampling using known variables such as habitat type, elevation, or soil type are recommended to be performed in both districts which will enhance the performance of all models.

## REFERENCES

Badia D., Ruiz A., Giona A., Marti C., Casanova J., Ibbara P., Zufiaurre R. (2016): The influence of elevation on soil properties and forest litter in the Siliceous Moncayo Massif, SW Europe. Journal of Mountain Science, 13: 2155–2169.

Borůvka L., Vašát R., Šrámek V., Neudertová Hellebrandová K., Fadrhonsová V., Sáňka M., Pavlů L., Sáňka O., Vacek O., Němeček K., Nozari S., Oppong Sarkodie V.Y. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil and Water Research, 17: 69–79.

Breiman L. (2001): Random forests. Machine Learning, 45: 5–32.

Brevik E.C. (2013): Soil health and productivity. In: Verheye W.H. (ed.): Soils, Plant Growth and Crop Production. Oxford, Encyclopedia of Life Support Systems (EOLSS).

Brewer S., Liaw A., Wiener M., Liaw M.A. (2015): Package Random Forest. Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4.6-14. Available on https://cran.r-project.org/web/packages/randomForest/index.html

Brungard C., Boettinger J.L., Duniway M.C., Wills S.A., Edwards T.C. (2015): Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239: 68–83.

Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. (2015):

https://doi.org/10.17221/119/2023-SWR

System for Automated Geoscientific Analyses (SAGA) V. 2.1.4. Geoscientific Model Development, 8: 1991–2007.

EEA (2018): CORINE Land Cover (CLC) 2018, Version 20, European Environment Agency, Kopenhagen, Denmark. Available on https://www.eea.europa.eu/data-and-maps/data/corine-land-cover-accounting-layers (accessed Jan 23, 2020).

Ellili Y., Walter C., Michot D., Pichelin P., Lemercier B. (2019): Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. Geoderma, 351: 1–8.

Fathololoumi S., Vaezi A.R., Alavipanah S.K., Ghorbani A., Saurette D., Biswas A. (2020): Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. Science of the Total Environment, 721: 137703.

Henderson B.L., Bui E.N., Moran C.J., Simon D.A.P. (2005): Australia-wide predictions of soil properties using decision trees. Geoderma, 124: 383–398.

Heung B., Bulmer C.E., Schmidt M.G. (2014): Predictive soil parent material mapping at a regional scale: A random forest approach. Geoderma, 214: 141–154.

IUSS Working Group WRB (2015): World Reference Base for Soil Resources 2014 (Update 2015). International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. World Soil Resources Reports No. 106, Rome, FAO.

James G., Witten D., Hastie T., Tibshirani R. (2014): An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Inc.

Kozák J., Němeček J., Jetmar M. (1996): The database of soil information system – PUGIS. Plant Production, 42: 529–534.

Kuhn M. (2012): The Caret Package. R Found. Statistical Computing. Vienna. Available on https//cran. r-project. org/packa-ge= caret.

Kuhn M., Weston S., Keefer C., Coulter N., Quinlan R. (2013): Cubist: Rule-and instance-based regression modeling. R Package Version 0.0.13. Wien, CRAN.

Lagacherie P., McBratney A.B., Voltz M. (eds.) (2006): Digital Soil Mapping – An Introductory Perspective. Developments in Soil Science, Vol. 31, Amsterdam, Hardbound, Elsevier Science.

Lal R., Bouma J., Brevik E., Dawson L., Field D.J., Glaser B., Hatano R., Hartemink A.E., Kosaki T., Lascelles B., Monger C., Muggler C., Martial Ndzana G., Norra S., Pan X., Paradelo R., Reyes-Sánchez L.B., Sandén T., Singh B.R., Spiegel H., Yanai J., Zhang J. (2021): Soils and sustainable development goals of the United Nations: An International Union of Soil Sciences perspective. Geoderma Regional, 25: e00398.

Lamichhane S., Kumar L., Wilson B. (2019): Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma, 352: 395–413.

Li L., Lu J., Wang S,. Ma Y,. Wei Q,. Li X,. Cong R., Ren T. (2016): Methods for estimating leaf nitrogen concentration of winter oilseed rape (*Brassica napus* L.) using in situ leaf spectroscopy. Industrial Crops and Products, 91: 194–204.

Lorenzetti R., Barbetti R., Fantappiè M., L'Abate G., Costantini E.A.C. (2015): Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small-scale maps. Geoderma, 237–238: 237–245.

McBratney A.B., Mendonça Santos M.L., Minasny B. (2003): On digital soil mapping. Geoderma, 117: 3–52.

Meinshausen N. (2006): Quantile regression forests. Journal of Machine Learning Research, 7: 983–999.

Meinshausen N., Hauser A., Mooij J., Peters J., Versteeg P., Buhlmann P. (2016): Methods for causal inference from gene perturbation experiments and validation. PNAS, 113: 7361–7368.

Miko L., Hošek M. (2009): The State of Nature and the Landscape in the Czech Republic. 1st Ed. Prague, Agency for Nature Conservation and Landscape Protection of the Czech Republic.

Minasny B., McBratney A.B. (2016): Digital soil mapping: A brief history and some lessons. Geoderma, 264: 301–311.

Minasny B., McBratney A.B., Salvador-Blanes S. (2008): Quantitative models for pedogenesis – A review. Geoderma, 144: 140–157.

Moore D.S., Notz W.I., Flinger M.A. (2013): The Basic Practice of Statistics. 6th Ed. New York, W.H. Freeman and Company.

Mosleh Z., Salehi M.H., Jafari A., Borujeni I.E., Mehnatkesh A. (2016): The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environmental Monitoring and Assessment, 188: 1–13.

Mulder V.L., de Bruin S., Schaepman M.E., Mayr T.R. (2011): The use of remote sensing in soil and terrain mapping – A review. Geoderma, 162: 1–19.

Němeček J. (2000): The status of soil mapping in the Czech Republic. The European Soil Information System. World Soil Resources Reports, 91: 61–65.

Němeček J., Mühlhanselová M., Macků J., Vokoun J., Vavříček D., Novák P. (2011): Taxonomical classification system of soils of the Czech Republic. 2nd modified Ed. Prague, ČZU, Prague. (in Czech)

Nikou M., Tziachris P. (2022): Prediction and uncertainty capabilities of quantile regression forests in estimating spatial distribution of soil organic matter. ISPRS International Journal of Geo-Information, 11: 130.

https://doi.org/10.17221/119/2023-SWR

Nozari S., Borůvka L. (2020): The effect of landscape slope on soil organic carbon in the Liberec district in the Czech Republic. International Journal of Advanced Science and Technology, 29: 3934–3942.

Nozari S., Borůvka L. (2023): The effects of slope and altitude on soil organic carbon and clay content in different land-uses: A case study in the Czech Republic. Soil and Water Research, 18: 204–218.

Pahlavan-Rad M.R., Akbari Moghaddam A.R. (2018): Spatial variability of soil texture fractions and pH in a flood plain (A case study from eastern Iran). Catena, 160: 275–281.

Pahlavan-Rad M.R., Dahmardeh K.H., Brungard C. (2018): Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. Geoderma Regional, 15: e00195.

Pahlavan-Rad M.R., Dahmardeh K.H., Hadizadeh M., Keykha G., Mohammadnia N., Gangali M., Keikha M., Davatgare N., Brungard C. (2020): Prediction of soil water infiltration using multiple linear regression and random forest in a dry flod plain, eastern Iran. Catena, 194: 104715.

Panagos P., Meusburger K., Ballabio C., Borrelli P., Alewell C. (2014): Soil erodibility in Europe: A high-resolution dataset based on LUCAS. The Science of the Total Environment, 479: 189–200.

Peters J., De Baets B., Verhoest N.E., Samson R., Degroeve S., De Becker P. Huybrechts W. (2007): Random forests as a tool for ecohydrological distribution modeling. Ecological Modeling, 207: 304–318.

Pieri C.J. (1992): The effect of population growth on the soil. In: Fertility of Soils. Springer Series in Physical Environment, Vol. 10, Berlin, Heidelberg, Springer: 84–99.

Pospíšil F. (1964): Fractionation of humus substances of several soil types in Czechoslovakia. Rostlinná Výroba, 10: 567–579.

Prasad A.M., Iverson L.R., Liaw A. (2006): Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 9: 181–199.

Pouladi N., Møller A.B. Tabatabai S., Greve M.H. (2019): Mapping soil organic matter contents at field level with cubist, random forest and kriging, Geoderma, 342: 85–92.

Quinlan J. (1993): C4.5: Programs for Empirical Learning. San Francisco, Morgan Kaufmann Publishers, Inc.

R Core Team (2018): R, a Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing.

Selvaradjou S.K., Montanarella L., Carre F., Jones A., Panagos P., Ragunath K., Kumaraperumal R., Natarajan S. (2007): An Innovative Approach for Updating Soil Information based on Digital Soil Mapping Techniques. Official Publication of European Communities. EUR 22545 EN: 1–34.

SPSS (2001): Statistics for Windows, Version 11.0. Chicago, SPSS Inc.

Taghizadeh-Mehrjardi R., Nabiollahi K., Kerry R. (2016): Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh Region, Iran. Geoderma, 266: 98–110.

Tesfahunegn G.B., Tamene L., Vlek P.L.G. (2011): Catchment scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia. Soil and Tillage Research, 117: 124–139.

Tziachris P., Aschonitis V., Chatzistathis T. (2019): Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. Catena, 174: 206–216.

USGS (2021): US Geological Survey. Available on https://earthexplorer.usgs.gov/.

Vaysse K., Lagacherie P. (2017): Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma, 291: 55–64.

Victoria R., Banwart S., Black H., Ingram J., Joosten H., Milne E., Noellemeyer E. (2012): The Benefits of Soil Organic Carbon. Emerging Issues in Our Global Environment. Nairobi, UNEP Division of Early Warning and Assessment United Nations Environment.

Weier J., Herring D. (2000): Measuring Vegetation (NDVI & EVI). NASA. Available on https://earthobservatory.nasa.gov/Features/MeasuringVegetation/ (accessed Jan 2020).

Wiesmeier M., Urbanski L., Hobley E.; Lang B., von Lützow M., Marin-Spiotta E., van Wesemael B., Rabot E., Ließ M., Garcia-Franco N. (2019): Soil organic carbon storage as a key function of soils – A review of drivers and indicators at various scales. Geoderma, 333: 149–162.

Winowiecki L., Vågen T.G., Massawe B. Jelinski N., Yamchai Ch., Sayula G., Msoka E. (2016): Landscape-scale variability of soil health indicators: effects of cultivation on soil organic carbon in the Usambara Mountains of Tanzania. Nutrient Cycling in the Agroecosystems, 105: 263–274.

Zeraatpisheh M., Ayoubi S., Jafari A., Tajik S., Finke P. (2019): Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma, 338: 445–452.

Zhu H., Hud W., Jing Y., Cao Y. (2018): Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform, Geoderma, 330: 9–18.

Žížala D., Minařík R., Skála J., Beitlerová H., Juřicová A., Rojas J.R., Penížek V., Zádorová T. (2022): High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena, 212: 106024.

Received: December 12, 2023
Accepted: January 3, 2024
Published online: January 15, 2024